




# Zoobot: Adaptable Deep Learning Models for Galaxy Morphology

Mike Walmsley <sup>1¶</sup>, Campbell Allen<sup>2\*</sup>, Ben Aussel <sup>3\*</sup>, Micah Bowles <sup>1\*</sup>, Kasia Gregorowicz <sup>4\*</sup>, Inigo Val Slijepcevic <sup>1\*</sup>, Chris J. Lintott <sup>5</sup>, Anna M. M. Scaife <sup>1,6</sup>, Maja Jabłońska <sup>4</sup>, Kosio Karchev <sup>7</sup>, Denise Lanzieri <sup>8</sup>, Devina Mohan <sup>1</sup>, David O’Ryan <sup>9</sup>, Bharath Saiguhan <sup>10</sup>, Crisel Suárez <sup>11,12</sup>, Nicolás Guerra-Varas <sup>13,14</sup>, and Renuka Velu<sup>15</sup>

**1** Jodrell Bank Centre for Astrophysics, Department of Physics and Astronomy, University of Manchester, Manchester, UK **2** Zooniverse.org, University of Oxford, Oxford, UK **3** Institut für Planetologie, Westfälische Wilhelms-Universität Münster, Münster, Germany **4** Astronomical Observatory of the University of Warsaw, Warsaw, Poland **5** Oxford Astrophysics, Department of Physics, University of Oxford, Oxford, UK **6** The Alan Turing Institute, London, UK **7** Theoretical and Scientific Data Science Group, Scuola Internazionale Superiore di Studi Avanzati (SISSA), Trieste Italy **8** Université Paris Cité, Université Paris-Saclay, CEA, CNRS, AIM, Gif-sur-Yvette, France **9** Department of Physics, Lancaster University, Lancaster, UK **10** Physical Research Laboratory, Navrangpura, Ahmedabad, India **11** Vanderbilt University, Nashville, USA **12** Center for Astrophysics | Harvard & Smithsonian, Cambridge, USA **13** Dipartimento di Fisica, Università di Roma “Tor Vergata”, Roma, Italy **14** Department of Astronomy, Faculty of Mathematics, University of Belgrade, Belgrade, Serbia **15** Ruprecht Karl University of Heidelberg, Germany ¶ Corresponding author \* These authors contributed equally.

DOI: [10.21105/joss.05312](https://doi.org/10.21105/joss.05312)

## Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: [Paul La Plante](#) 

## Reviewers:

- [@crhea93](#)
- [@devanshkv](#)

Submitted: 09 March 2023

Published: 08 May 2023

## License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

## Summary

Zoobot is a Python package for measuring the detailed appearance of galaxies in telescope images using deep learning. Zoobot is aimed at astronomers who want to solve a galaxy image task such as finding merging galaxies or counting spiral arms. Astronomers can use Zoobot to adapt (finetune) pretrained deep learning models to solve their task. These finetuned models perform better and require far fewer new labels than training from scratch ([Walmsley, Slijepcevic, et al., 2022](#)).

The models included with Zoobot are pretrained on up to 92 million responses from [Galaxy Zoo](#) volunteers. Each volunteer answers a series of tasks describing the detailed appearance of each galaxy. Zoobot’s models are trained to answer all of these diverse tasks simultaneously. The models can then be adapted to new related tasks.

Zoobot provides a high-level API and guided workflow for carrying out the finetuning process. The API abstracts away engineering details such as efficiently loading astronomical images, multi-GPU training, iteratively finetuning deeper model layers, and so forth. Behind the scenes, these steps are implemented via either PyTorch or TensorFlow, according to the user’s choice. Zoobot is therefore accessible to astronomers with no previous experience in deep learning.

For advanced users, Zoobot also includes the code to replicate and extend our pretrained models. This is used routinely at [Galaxy Zoo](#) to scale up galaxy measurement catalogs ([Walmsley, Lintott, et al., 2022](#)) and to prioritise the galaxies shown to volunteers for labelling. Zoobot models have been applied to measure galaxy appearance in SDSS ([Walmsley et al., 2020](#)), Hubble, HSC, and DESI, and are included in the data pipeline of upcoming space telescope Euclid ([Laureijs et al., 2011](#)). We hope that Zoobot will help empower astronomers to apply deep learning to answer their own science questions.

## Statement of need

One common way to investigate why galaxies look the way they do is by measuring the appearance - morphology - of millions of galaxies and looking for connections between appearance and other physical properties (Masters, 2019). The sheer number of images requires most of these measurements to be made automatically with software (Walmsley et al., 2020).

Unfortunately, making automated measurements of complicated features like spiral arms is difficult because it is hard to write down a set of steps that reliably identify those and only those features. This mirrors many image classification problems back on Earth (LeCun et al., 2015). Astronomers often aim instead to learn the measurement steps directly from data by providing deep learning models with large sets of galaxy images with labels (e.g. spiral or not) (Huertas-Company & Lanusse, 2022).

Gathering large sets of labelled galaxy images is a major practical barrier. Models trained on millions to billions of labelled images consistently perform better (Bommasani et al., 2021; Dehghani et al., 2023), but astronomers cannot routinely label this many images. Neither can most other people; terrestrial practitioners often start with a model already trained (“pretrained”) on a broad generic task and then adapt it (“finetune”) to their specific measurement task (Ridnik et al., 2021).

Zoobot makes this approach available to astronomers. We provide models pretrained on millions of galaxy images and present a convenient API for finetuning those models. Early results (O’Ryan et al., 2023; Walmsley, Slijepcevic, et al., 2022) show that our pretrained models can be efficiently adapted to new morphology tasks.

## Acknowledgements

Zoobot was made possible by the Galaxy Zoo volunteers, who collectively created the crucial morphology labels used to create our pretrained models (and much, much more). Their efforts are individually and gratefully acknowledged [here](#). Thank you.

MW, IVS, MB and AMS gratefully acknowledge support from the UK Alan Turing Institute under grant reference EP/V030302/1. IVS gratefully acknowledges support from the Frankopan Foundation. CS acknowledges the support NASA Heliosphysics Supporting Research Grant (NASA-HSR 16611153) and Vanderbilt Bridge Program PhD Fellowship. DOR acknowledges the support of the UK Science and Technology Facilities Council (STFC) under grant reference ST/T506205/1. BS gratefully acknowledges support from the Physical Research Laboratory. We thank Astro Hack Week for bringing together a community to experiment with the beta version of Zoobot.

Our (recommended) PyTorch (Paszke et al., 2019) version uses timm (Wightman, 2019) for architecture definition, Pyro (Bingham et al., 2019; Phan et al., 2019) for our custom loss and Pytorch Lightning (Falcon & The PyTorch Lightning team, 2019) for training. Our TensorFlow (Abadi et al., 2016) version uses Tensorflow Probability (Dillon et al., 2017) for our custom loss.

For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising.

## References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., ... Zheng, X. (2016). *TensorFlow: Large-scale*

- machine learning on heterogeneous distributed systems*. arXiv. <https://doi.org/10.48550/ARXIV.1603.04467>
- Bingham, E., Chen, J. P., Jankowiak, M., Obermeyer, F., Pradhan, N., Karaletsos, T., Singh, R., Szerlip, P. A., Horsfall, P., & Goodman, N. D. (2019). Pyro: Deep universal probabilistic programming. *J. Mach. Learn. Res.*, 20, 28:1–28:6. <http://jmlr.org/papers/v20/18-403.html>
- Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., Arx, S. von, Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., ... Liang, P. (2021). On the opportunities and risks of foundation models. In *arXiv preprint*. <https://doi.org/10.48550/arXiv.2108.07258>
- Dehghani, M., Djolonga, J., Mustafa, B., Padlewski, P., Heek, J., Gilmer, J., Steiner, A., Caron, M., Geirhos, R., Alabdulmohsin, I., Jenatton, R., Beyer, L., Tschannen, M., Arnab, A., Wang, X., Riquelme, C., Minderer, M., Puigcerver, J., Evci, U., ... Houlsby, N. (2023). *Scaling vision transformers to 22 billion parameters*. arXiv. <https://doi.org/10.48550/ARXIV.2302.05442>
- Dillon, J. V., Langmore, I., Tran, D., Brevdo, E., Vasudevan, S., Moore, D., Patton, B., Alemi, A., Hoffman, M., & Saurous, R. A. (2017). *TensorFlow distributions*. arXiv. <https://doi.org/10.48550/ARXIV.1711.10604>
- Falcon, W., & The PyTorch Lightning team. (2019). *PyTorch Lightning* (Version 1.4). <https://doi.org/10.5281/zenodo.3828935>
- Huertas-Company, M., & Lanusse, F. (2022). The DAWES review 10: The impact of deep learning for the analysis of galaxy surveys. *Publications of the Astronomical Society of Australia*. <https://doi.org/10.1017/pasa.2022.55>
- Laureijs, R., Amiaux, J., Arduini, S., Auguères, J.-L., Brinchmann, J., Cole, R., Cropper, M., Dabin, C., Duvet, L., Ealet, A., Garilli, B., Gondoin, P., Guzzo, L., Hoar, J., Hoekstra, H., Holmes, R., Kitching, T., Maciaszek, T., Mellier, Y., ... Zucca, E. (2011). Euclid Definition Study Report. *arXiv e-Prints*, arXiv:1110.3193. <https://doi.org/10.48550/arXiv.1110.3193>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521, 436–444. <https://doi.org/10.1038/nature14539>
- Masters, K. L. (2019). Twelve years of galaxy zoo. *Proceedings of the International Astronomical Union*, 14, 205–212. <https://doi.org/10.1017/S1743921319008615>
- O’Ryan, D., Merín, B., Simmons, B. D., Vojteková, A., Anku, A., Walmsley, M., Garland, I. L., Géron, T., Keel, W., Kruk, S., Lintott, C. J., Mantha, K. B., Masters, K. L., Reerink, J., Smethurst, R. J., & Thorne, M. R. (2023). *Harnessing the hubble space telescope archives: A catalogue of 21,926 interacting galaxies*. arXiv. <https://doi.org/10.48550/ARXIV.2303.00366>
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., ... Chintala, S. (2019). PyTorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. dAlché-Buc, E. Fox, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems 32* (pp. 8024–8035). Curran Associates, Inc. <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>
- Phan, D., Pradhan, N., & Jankowiak, M. (2019). Composable effects for flexible and accelerated probabilistic programming in NumPyro. *arXiv Preprint arXiv:1912.11554*. <https://doi.org/10.48550/arXiv.1912.11554>

- Ridnik, T., Ben-Baruch, E., Noy, A., & Zelnik-Manor, L. (2021). *ImageNet-21K pretraining for the masses*. arXiv. <https://doi.org/10.48550/ARXIV.2104.10972>
- Walmsley, M., Lintott, C., Tobias, G., Kruk, S. J., Krawczyk, C., Willett, K., Bamford, S., Keel, W., Kelvin, L. S., Fortson, L., Masters, K., Mehta, V., Simmons, B., Smethurst, R. J., Baeten, E. M. L., & Macmillan, C. (2022). Galaxy zoo DECaLS: Detailed visual morphology measurements from volunteers and deep learning for 314,000 galaxies. *Monthly Notices of the Royal Astronomical Society*, 509, 3966–3988. <https://doi.org/10.1093/mnras/stab2093>
- Walmsley, M., Slijepcevic, I. V., Bowles, M., & Scaife, A. M. M. (2022, June). Towards galaxy foundation models with hybrid contrastive learning. *Machine Learning for Astrophysics Workshop at the Thirty Ninth International Conference on Machine Learning (ICML 2022)*. <https://doi.org/10.48550/arxiv.2206.11927>
- Walmsley, M., Smith, L., Lintott, C., Gal, Y., Bamford, S., Dickinson, H., Fortson, L., Kruk, S., Masters, K., Scarlata, C., Simmons, B., Smethurst, R., & Wright, D. (2020). Galaxy zoo: Probabilistic morphology through bayesian CNNs and active learning. *Monthly Notices of the Royal Astronomical Society*, 491, 1554–1574. <https://doi.org/10.1093/mnras/stz2816>
- Wightman, R. (2019). PyTorch image models. In *GitHub repository*. <https://github.com/rwightman/pytorch-image-models>; GitHub. <https://doi.org/10.5281/zenodo.4414861>